Al-Nahrain Journal for Engineering Sciences **NJES** 24(2)144-148, 2021
http://doi.org/10.29194/NJES.24020144

Facial Expression Recognition Based on Texture Features

Alaa Nabeel Haj-Najeb[1], Nasser Ali Nasser[2]

Authors affiliations:
1) Software and Information Systems Dept., Tishreen University, Latakia, Syrian Arab Republic.
alaanajeeb91@gmail.com

2) Software and Information Systems Dept., Tishreen University, Latakia, Syrian Arab Republic.
n.a.nasser@gmail.com

Paper History:

Received: 5th Aug. 2021

Revised: 5th Oct. 2021

Accepted: 20th Oct. 2021

Abstract

Facial expressions are a form of non-verbal communication, they appear as changes on the surface of the facial skin according to one's inner emotional states, aims, or social communications. Classification of these expressions is a normal process for humans, but it is a challenging task for machines.

Lately, interest in facial expression recognition has grown, and many systems have been developed to classify expressions from facial images. Any expression recognition system is comprised of three steps. The first one is face acquisition, then feature extraction, and finally classification. The classification accuracy depends primarily on the feature extraction step. Therefore, in this research we study many texture feature extraction descriptors and compare their results under the same preprocessing circumstances; moreover, we propose two improvements for one of these descriptors, which give better results than the original one. We validate the results on two commonly used databases for expression recognition using Matlab programming language, wishing all of that to be an interesting point for researchers in this field.

Keywords: Expression Recognition, Feature Extraction, Texture, LBP Variants, Image Processing.

التعرف على تعابير الوجه بناءً على سمات النسيج

آلاء نبيل حاج نجيب، ناصر علي ناصر

الخلاصة:

تعابير الوجه هي شكل من أشكال التواصل الغير لفظي، تظهر هذه التعابير كتغيرات على سطح بشره الوجه نتيجة حالة الشخص العاطفية الداخلية او نواياه او نتيجة التواصل الاجتماعي. تمييز هذه التعابير يتم بشكل طبيعي عند البشر لكن الامر ليس بهذه البساطة بالنسبة للآلة.

مع ازدياد الاهتمام بهذا المجال مؤخراً، تم تطوير الكثير من الأنظمة لتمييز التعابير من صور الوجه، يتكون أي نظام من ثلاث مراحل تبدأ بتحديد الوجه، ثم استخلاص السمات وبعدها مرحلة التصنيف. تعتمد دقة التمييز لهذه الأنظمة بشكل أساسي على قوة خوارزمية استخراج السمات لذلك قمنا في هذا البحث بدراسة خوارزميات استخراج السمات عن طريق القوام ومقارنة نتائجهم باستخدام نفس شروط المعالجة المُسبقة للصور، بالإضافة لوضع اقتراحين لتحسين احدى هذه الخوارزميات، وكانت نتائج كلا التحسينين أفضل من نتائج الخوارزمية الأساسية. تم التحقق من النتائج باستخدام قاعدتي بيانات تُستخدمان لتمييز التعابير وذلك بالاعتماد على لغة البرمجة ماتلاب، آملين ان يكون هذا البحث نقطة اهتمام للباحثين في هذا المجال.

1. Introduction

Facial expressions play a fundamental role in social communication allowing members of a community to express themselves by using non-verbal cues and thereby conveying important information. Professor Mehrabian shows that 7% of the message meaning is communicated through spoken word, 38% through tone of voice, and 55% through facial expression [1]. Therefore, this field has attracted the researchers' interest, so they started with studying these expressions and observing their diversity across cultures, focusing on the similarities as well as the differences on how they were expressed and received. Paul Ekman concluded that all facial expressions are universal among humans regardless of their age, race and gender [2]. In spite of the variety of facial expressions, the main focus was only on six basic expressions which are anger, disgust, fear, sadness, happiness, and surprise. Automatic facial expression can be applied in a wide range of areas, such as psychology, psychiatry, neurology, pain assessment, lie detection, human computer interface [3]. Moreover, it is one of the challenges in face recognition field. [4]

Timo Ojala et al [5] introduced LBP operator for texture analysis, then it has been effectively utilized as a local feature extractor in facial expression recognition. It encodes the local texture of an image by quantizing the $\boldsymbol{P}$ neighbor gray levels of a local neighborhood with respect to the center pixel. LBP is computationally effective and robust to monotonic intensity variations; however, it degrades under the existence of large illumination variation and random noise, since a little modification in the gray level value can simply alter the LBP code. To address these issues and improve its performance, many variants of LBP have been developed. Taskeed Jabid et al [6] handled LBP problems by calculating the edge response values in eight directions using Kirsch masks at each pixel and encoding them into an eight bit binary number using the relative strength of these edge responses, this method is called Local Directional Pattern (LDP). LDP is more robust than LBP against noise and non-monotonic illumination changes because it depends on gradients, which are more stable than gray values. It also describes the local primitives such as curves, corners, and junctions more stably and holds more information. Hasanul Kabir et al [7] started with the idea that texture can be well represented when described by a spatial structure in addition to its contrast. However, LDP does not contain information about contrast. Therefore, they included contrast in the feature vector by adding the variance as an adaptive weight to adjust the contribution of the LDP code in the histogram formation, and they called it Local Directional Pattern Variance (LDPv). Faisal Ahmed et al [8] found that LBP encodes only the signs of the gray levels' differences. Thus, it discards some main texture information. Therefore, to improve the performance, they presented Compound Local Binary Pattern (CLBP) which adds extra $\boldsymbol{P}$ bits with the native LBP code to exploit both the sign and the magnitude information of the differences between the center and the neighbor gray values. Ramirez Rivera et al [9] found that LDP misses some directional information (the responses' sign) by treating all directions in the same way; also, it still suffers in non-monotonic illumination variation and random noise. Therefore, they presented Local Directional Number Pattern (LDN) that represents the structure of the texture and its intensity transitions. They computed the edge responses of the neighborhood using compass masks, which extract directional information and encoded such information using the outstanding direction indexes and sign to discriminate among similar structural patterns that have different intensity transitions (e.g., from bright to dark and vice versa) in the texture producing a more discriminative code than current methods. Wenjin Chu et al [10] proposed Gradient Directional Pattern (GDP) operator, which utilizes the more stable gradient direction values instead of gray levels. The texture-coding scheme quantizes the gradient direction angle in a local neighborhood according to the center pixel gradient angle and a threshold. The threshold facilitates the generation of consistent texture patterns in smooth facial regions. Ying Tong et al [11] saw that LBP does not precisely describe the texture of the facial muscles and other local distortion because it depends on comparing the gray level value of the center pixel and neighboring pixels. Therefore, they proposed Local Gradient Coding (LGC) algorithm, which compares the eight peripheral pixels, vertical, horizontal, and diagonal gradients. In order to minimize the computational complexity and omit the redundant while not miss the main information in the face texture expression, they proposed and optimized a new LGC operator based only on horizontal and diagonal gradient and called it LGC based on Horizontal and Diagonal prior principle.(LGC-HD). Jucheng Yang et al [12] found that there is still a necessity for upgrading the accuracy of other feature extraction methods while preserving their speed and low computational complexity. Therefore, they improved LGC-HD, which has performance limitations on large-scale problems because it works on 3*3 neighborhood. Hence, they proposed Central Symmetric Local Gradient Coding (CS-LGC) which works on 5*5 grids and uses four different directional gradients of the horizontal, vertical, and diagonal lines. The gradient is computed in a center-symmetric way so that the resulted feature values of the target pixels are more representative.

2. Proposed improvements

We select GDP operator because it has a room for improvement. For example, GDP uses a threshold to generate binary code; this threshold may vary from one database to another. We try to combine GDP with other operators to improve its results under the same used circumstances; for instance, image size, number of sub regions, classifier, and evaluation technique.

2.1. Local Gradient Directional Number (LGDN)

Like GDP operator, we start by calculating Sobel filter response values. This filter convolves the image with a horizontal and a vertical kernel to get the values of horizontal (Gx) and vertical (Gy) gradients. Sobel filter is shown in Fig. 1.

−1	−2	−1
0	0	0
1	2	1

(a)

−1	0	1
−2	0	2
−1	0	1

(b)

Figure (1): Sobel operator (a): horizontal filter (b): vertical filter.

Then the gradient direction of each pixel in an image is computed using eq.(1)

$$a(x,y) = tan^{-1}\,(Gy/\,Gx) \qquad ...(1)$$

After that, we compare the eight neighboring angles of each pixel, and include directional information by encoding the prominent direction indexes and sign of maximum positive and minimum negative angle. The maximum positive angle indicates a vertical edge, and the minimum negative angle (which is another maximum in absolute value) indicates a vertical edge with opposite direction in comparison with the previous case. Then, we set the index of maximum positive angle to the most three significant bits, and the index of minimum negative angle to the least three significant bits. We convert the binary code into decimal one to get encoded image. Finally, we calculate histogram and get feature vector of length 56 instead of 256 (as in GDP) because we have six bits for each pixel. Fig.2 shows LGDN code generation. To sum up, we get benefit of an idea from LDN to improve GDP.

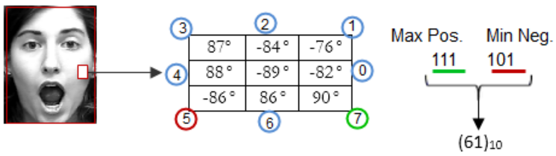
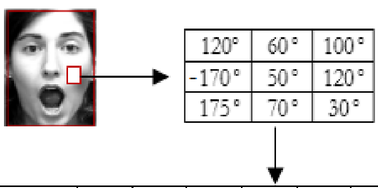


Figure (2): Generating LGDN code

2.2. Local Gradient Directional Pattern (LGDP)

Like GDP operator and previous improvement, we calculate Sobel filter responses and gradient direction angle of each pixel using eq.(1) . Then we compare the eight gradient direction angles for each pixel and include information about the importance of these responses -as high responses indicate the presence of corner or vertical edge- by setting the top absolute k values to one, and the other absolute (8-k) to zero. We choose k=3 because it gives the best results. Then we convert the binary code into decimal one to get encoded image. Finally, we calculate histogram from the encoded image and get feature vector of length (C^8_3= 56) instead of 256 because we have six bits for each pixel. Fig.3 shows LGDP code generation. To sum up, we get benefit of an idea from LDP to improve GDP.



Neighbor index	0	1	2	3	4	5	6	7
Angle value	120	100	60	120	-170	175	70	30
Rank	3	5	7	4	2	1	6	8
Code bit	1	0	0	0	1	1	0	0
Final decimal code: 49								

Figure (3): Generating LGDP code

3. Experimental setup

In this work, we evaluate the previous descriptors on two commonly used databases with seven expressions (six basic and natural one). The first one is Japanese Female Facial Expression (JAFFE) database; it contains 213 images of facial expressions for ten female subjects. The resolution of each image is 256* 256 pixels. We extract faces for 199 images, then we cut and resize all the images to 110 * 150 pixels. The second one is Cohn-Kanade (CK) database, which comprises 100 university students who performed a sequence of 23 facial displays, six of them were based on six basic emotions. Image sequences from neutral to target expression were captured. In our experiments, we select the three most expressive image frames for basic expressions. Then, we add the first image (neutral expression) from the same sequences to form the 7-class expression data set. At last, we get 375 images from 20 subjects. Next, we cut and resize all the images to 175 * 195 pixels.

We apply each descriptor on these databases to get encoded images. Then, we calculate histogram, which can effectively represent the global features of image texture. Histogram generated from each image describes the statistical distribution of the local micro-structures; however, it doesn't contain spatial information of such distribution. Therefore, each image is split into *m*n* non-overlapping regions and the histogram is extracted for each region then these histograms are concatenated to get the final feature vector as the facial descriptor. The previous setup steps are summarized in Fig.4.

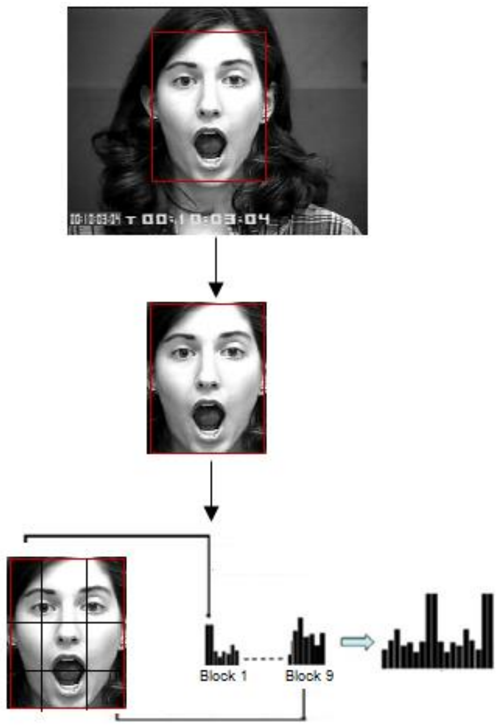


Figure (4): System architecture

For classification step, we adopt SVM classifier to classify facial expression. To evaluate the performance, we randomly divide each database into 10 sections and leave-one-out cross validation was conducted.

4. Experimental Results and Discussion

4.1. Influence of different block manner:

We pick out three different divisions: 3*3, 5*5 and 7*6 sub-regions to evaluate different partitioning methods. The recognition rates of the tested algorithms (called by their acronym) and our proposed modifications are listed in Table 1 and Table 2.

Table (1): Recognition rates **(%)** of the descriptors using several subdivisions on CK database. The symbol (-) indicates that results are not available owing to the doubling size of the feature vector in the used environment.

	3*3	**5*5**	**7*6**
LBP	75.20	**82.93**	81.86
CLBP	85.60	**91.46**	-
LDP	63.46	65.33	**69.86**
LDPv	66.13	78.13	**80**
LDN	79.20	90.66	**91.20**
GDP	68.80	77.60	**80**
LGC-HD	90.13	**96.26**	95.73
CS-LGC	76.80	84.80	**85.33**
LGDN	83.73	90.13	**93.60**
LGDP	74.93	83.73	**88.80**

Table (2): Recognition rates (%) of the descriptors using various subdivisions on JAFFE database.

	3*3	**5*5**	**7*6**
LBP	53.76	**59.79**	55.27
CLBP	64.82	**67.33**	65.82
LDP	63.31	66.33	**66.83**
LDPv	64.82	**68.34**	67.83
LDN	59.29	**72.86**	70.85
GDP	45.72	53.76	**55.27**
LGC-HD	66.33	**78.39**	74.37
CS-LGC	55.77	**61.30**	58.29
LGDN	57.28	61.30	**66.83**
LGDP	51.25	**60.80**	58.79

From recognition rates, we can conclude that:

1. It is not necessary for a particular descriptor to work well for both CK and JAFFE databases. For example, LDP works better than many descriptors on JAFFE database, but it gives lower results than other algorithms on CK database.
2. It is not necessary for a particular partitioning of a descriptor to work well for both CK and JAFFE databases. For example, LDPv and LDN give the best results for 5*5 sub regions on JAFFE database whereas they give best results for 7*6 on CK database.
3. Results on CK database are **better** than JAFFE database due to many reasons:
   - We use 199 images from JAFFE database, whereas 375 images from CK database, which means more features used from CK to train the classifier and increase recognition accuracy.
   - A single expression can be expressed in more than one way in JAFFE database. For example, these three expressions in Fig.5. are labeled as Anger in JAFFE database.

Figure (5): Different ways of expressing Anger

4. Recognition rates of both GDP proposed modifications are better than original GDP.
5. Feature vector length of each GDP proposed modification is shorter than the original GDP feature vector, so they are suitable for large databases.
6. It is possible to improve the recognition rate of an existing descriptor by making use of the strength point of another one.

4.2. Influence of illumination variation preprocessing:

Both of CK and JAFFE databases suffer from illumination variation. In Fig.6 we show examples from JAFFE database where the same person is affected by illumination across expressions. Moreover, we show examples from CK database where the same expression is affected by illumination changes across people.

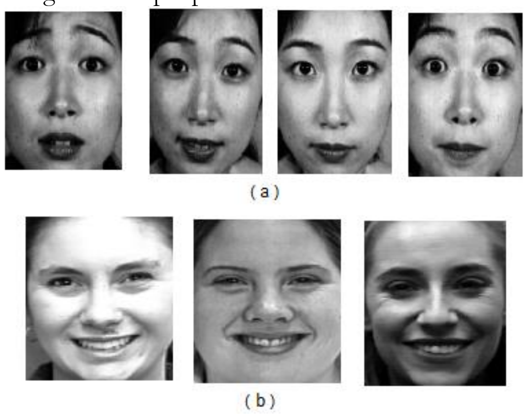


Figure (6): illumination variation (a): JAFFE database (b): CK database.

In order to eliminate the effect of the illumination variation, we adopt histogram equalization. It uses cumulative density function of the image then adjusts the brightness of an image by flattening the histogram to give it a linear trend, so the intensity values of the image will be spread out over all grey levels. [13].

To test the efficiency of histogram equalization, we apply it as a preprocessing stage and retest the descriptors on CK and JAFFE databases. Table 3 shows recognition rates for 5*5 and 7*6 sub regions (which give best results in previous experiment). We compare our results in this experiment with last experiment and conclude:

1. Histogram equalization improves the results because it removes illumination changes.
2. For LGC-HD operator, the results get fewer with histogram equalization for 5*5 sub regions because of this partitioning, while they are still

better with histogram equalization for 7*6 sub regions. So the setup has an important effect on recognition accuracy.

Table (3): Recognition rates (%) of the descriptors after histogram equalization.

	CK		JAFFE	
	5*5	**7*6**	**5*5**	**7*6**
LBP	83.20	83.20	62.31	56.78
CLBP	94.66	--	68.84	68.34
LDP	77.86	85.86	70.85	72.36
LDPv	81.60	86.13	72.36	72.36
LDN	91.20	92.26	69.34	72.36
GDP	78.93	78.66	55.27	54.27
LGC-HD	95.46	97.06	74.87	77.88
CS-LGC	86.66	86.66	62.81	61.80
LGDN	91.46	93.33	63.81	69.34
LGDP	85.60	90.66	60.30	59.79

5. Conclusion:

This paper compares some of texture-based feature extraction operators, which have been utilized for facial expression recognition. A performance evaluation of the tested descriptors is conducted and presented once for original images and once after illumination variation removal. In total, 8 descriptors have been tested on two databases under the same experimental settings. Moreover, we present two modifications for GDP descriptor and they give better results than original GDP on CK and JAFFE databases. Comparing all mentioned descriptors, we can see that the best recognition rate under selected settings is for LGC-HD using 5*5 sub regions for CK and JAFFE database.

In the future work, we will improve the existing methodologies especially in the pre-processing step because it has an important effect on extracted feature and recognition accuracy. Also we will try to modify the operators to get shorter feature vector and accelerate recognition process. We may extract features using deep learning methods and compare them with traditional methods.

6 References:

[1] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial Expression Recognition: A Survey," Symmetry, vol. 11, no. 10, pp. 1189, 2019.

[2] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art," arXiv:1203.6722 , Mar. 2012.

[3] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial Expression Analysis," Handbook of Face Recognition, pp. 247–275, 2005

[4] L. Ibrahim, N. Nasser, and M. Ali, "Face recognition Based on Statistical Texture Features," Embedded Self Organizing Systems, vol. 7, no. 1, pp. 10–15, 2020.

[5] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," Image and Vision Computing, vol. 27, no. 6, pp. 803–816, 2009.

[6] T. Jabid, Md. H. Kabir, and O. Chae, "Facial expression recognition using Local Directional Pattern (LDP)," IEEE Xplore, pp. 1605-1608, 2010.

[7] Md. H. Kabir, T. Jabid, and O. Chae, "A Local Directional Pattern Variance (LDPv) Based Face Descriptor for Human Facial Expression Recognition," 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 526-532, 2010.

[8] F. Ahmed, E. Hossain, A. S. M. H. Bari, and A. Shihavuddin, "Compound local binary pattern (CLBP) for robust facial expression recognition," 2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI), pp.391-395, 2011.

[9] A. Ramirez Rivera, J. Rojas Castillo, and O. Oksam Chae, "Local Directional Number Pattern for Face Analysis: Face and Expression Recognition," IEEE Transactions on Image Processing, vol. 22, no. 5, pp. pp.1458-1462, 2013.

[10] W. Chu, Z. Ying, and X. Xia, "Facial Expression Recognition Based on Local Binary Pattern and Gradient Directional Pattern," 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp. 1458-1462, 2013.

[11] Y. Tong, R. Chen, and Y. Cheng, "Facial expression recognition algorithm using LGC based on horizontal and diagonal prior principle," Optik, vol. 125, no. 16, pp. 4186–4189, 2014.

[12] J. Yang, X. Wang, S. Han, J. Wang, D. S. Park, and Y. Wang, "Improved Real-Time Facial Expression Recognition Based on a Novel Balanced and Symmetric Local Gradient Coding," Sensors, vol. 19, no. 8, p. 1899, Apr. 2019.

[13] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," Procedia Computer Science, vol. 116, pp. 523–529, Jan. 2017